

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 2001-228893

(43)Date of publication of application : 24.08.2001

(51)Int.Cl. G10L 15/20
G10L 21/02
G10L 11/00

(21)Application number : 2000-041086

(71)Applicant : MATSUSHITA ELECTRIC IND CO LTD

(22)Date of filing : 18.02.2000

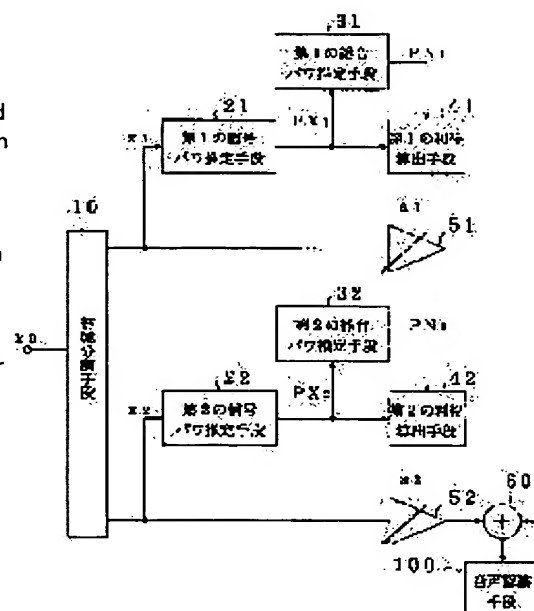
(72)Inventor : KANAMORI TAKEO
NAKATO YOSHIHISA
KAWAMURA TAKESHI

(54) SPEECH-RECOGNIZING DEVICE

(57)Abstract:

PROBLEM TO BE SOLVED: To realize a speech-recognizing device, which is capable of improving the speech recognition performance under a noisy environment by a low processing quantity.

SOLUTION: This device is provided with a band-dividing means 10 for dividing the band of input signals as fore-stage processing of a speech-recognizing means 100, by which the input signals are divided to bass and treble region sides. The signal power of the respective bands is estimated in first and second signal power estimating means 21 and 22, and noise power is estimated by first and second noise power estimating means 31 and 32. First and second gain calculating means 41 and 42 compute optimum gains and controls the amplification rates of first and second signal amplifier means 51 and 52. A signal adding means 60 adds the output signals of the signal amplifier means 51 and 52 and applies these signals to the speech-recognizing means 100. The distortion of the speech spectra by the noise can then be corrected by smaller number of the band divisions.



LEGAL STATUS

[Date of request for examination]

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's decision of rejection]

[Date of extinction of right]

(19)日本国特許庁 (J P)

(12) 公 開 特 許 公 報 (A)

(11)特許出願公開番号
特開2001-228893
(P2001-228893A)

(43)公開日 平成13年 8月24日 (2001. 8. 24)

(51)Int.Cl.	識別記号	F I	テ-リ-ト*(参考)
G 1 0 L	15/20	G 1 0 L	3 0 1 D
	21/02		5 D 0 1 5
	11/00		A
			9 A 0 0 1
			F

審査請求 未請求 請求項の数12 O L (全 13 頁)

(21)出願番号 特願2000-41086(P2000-41086)

(22)出願日 平成12年 2月18日 (2000. 2. 18)

(71)出願人 000005821

松下電器産業株式会社

大阪府門真市大字門真1006番地

(72)発明者 金森 丈郎

大阪府門真市大字門真1006番地 松下電器
産業株式会社内

(72)発明者 中藤 良久

大阪府門真市大字門真1006番地 松下電器
産業株式会社内

(74)代理人 100084364

弁理士 岡本 宜喜

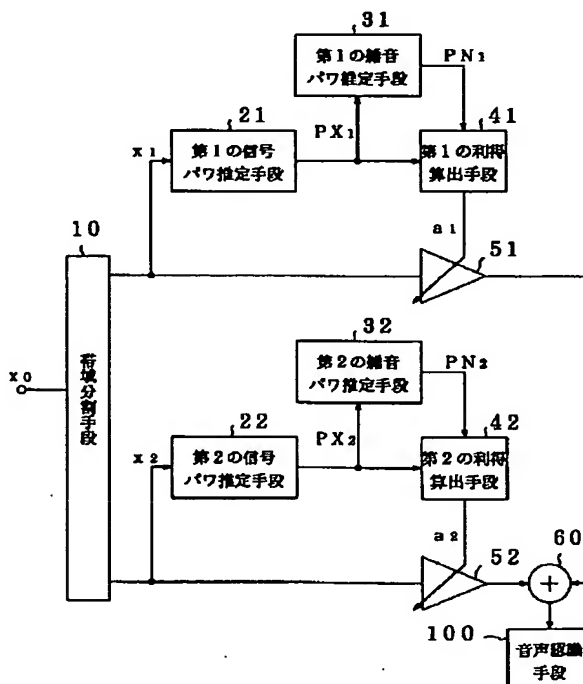
最終頁に続く

(54)【発明の名称】 音声認識装置

(57)【要約】

【課題】 雑音環境下における音声認識性能を、低処理量で改善できる音声認識装置を実現すること。

【解決手段】 音声認識手段100の前段処理として、入力信号を帯域分割する帯域分割手段10を設け、低域側と高域側に分割する。第1及び第2の信号パワ推定手段21、22で各帯域の信号パワを推定し、第1及び第2の雑音パワ推定手段31、32で雑音パワを推定する。第1及び第2の利得算出手段41、42は最適利得を演算し、第1及び第2の信号増幅手段51、52の増幅率を制御する。信号加算手段60は信号増幅手段51、52の出力信号を加算し、音声認識手段100に与える。こうすると、少ない帯域分割数で雑音による音声スペクトルの歪みを補正することができる。



【特許請求の範囲】

【請求項1】 入力音声に混入した雑音成分によって変形した音響信号に対して、信号スペクトルの形状を補正して音声認識を行う音声認識装置であって、入力音響信号を n 個の帯域に分割する帯域分割手段と、前記帯域分割手段によって n 分割された周波数帯域の音響信号を入力し、制御利得に基づいて音響信号の増幅又は減衰を行う第1～第 n の信号増幅手段と、前記帯域分割手段からの出力信号のレベルに応じて前記制御利得を演算し、前記第1～第 n の信号増幅手段の利得を制御する第1～第 n の利得制御手段と、前記第1～第 n の信号増幅手段からの出力信号を加算し、雑音成分の低減された音声信号を出力する信号加算手段と、前記信号加算手段から出力された音声信号を入力し、音声認識を行う音声認識手段と、を具備することを特徴とする音声認識装置。

【請求項2】 前記利得制御手段は、前記帯域分割手段で分割された帯域信号を入力し、短時間平均パワを出力する信号パワ推定手段と、前記信号パワ推定手段からの出力信号を入力し、信号の雑音パワを推定する雑音パワ推定手段と、前記信号パワ推定手段の出力信号と前記雑音パワ推定手段の出力信号とを入力し、入力信号パワと雑音パワの割合に応じた利得量となるように前記利得制御手段の制御利得を算出する利得算出手段と、を有し、前記利得算出手段は、制御利得を a_i とし、信号パワを PXi ($i=1\sim n$)とし、雑音パワを PNi とするとき、前記制御利得を $a_i = (PXi - PNi) / PXi$ により算出することを特徴とする請求項1記載の音声認識装置。

【請求項3】 前記利得制御手段は、前記帯域分割手段で分割された帯域信号を入力し、短時間平均パワを出力する信号パワ推定手段と、前記信号パワ推定手段からの出力信号を入力し、信号の雑音成分パワを推定する雑音パワ推定手段と、前記信号パワ推定手段の出力信号と前記雑音パワ推定手段の出力信号とを入力し、入力信号パワと雑音パワの割合に応じた利得量となるように前記利得制御手段の制御利得を算出する利得算出手段と、を有し、前記利得算出手段は、制御利得を a_i とし、信号パワを PXi ($i=1\sim n$)とし、雑音パワを PNi とし、利得制御範囲パラメータを Bi ($0 < Bi$) とするとき、前記制御利得を $a_i = (PXi - Bi \times PNi) / PXi$ により算出することを特徴とする請求項1記載の音声認識装置。

【請求項4】 前記利得制御手段は、前記帯域分割手段で分割された帯域信号を入力し、短時間平均パワを出力する信号パワ推定手段と、前記信号パワ推定手段からの出力信号を入力し、信号の

雑音パワを推定する雑音パワ推定手段と、前記信号パワ推定手段の出力信号と前記雑音パワ推定手段の出力信号とを入力し、入力信号パワと雑音パワの割合に応じた利得量となるように前記利得制御手段の制御利得を算出する利得算出手段と、を有し、前記利得算出手段は、制御利得を a_i とし、信号パワを PXi ($i=1\sim n$)とし、雑音パワを PNi とし、利得制御範囲パラメータを Bi ($0 < Bi$)、 Ci ($0 < Ci$) とするとき、前記制御利得を $a_i = (PXi - Bi \times PNi) / PXi + Ci$ により算出することを特徴とする請求項1記載の音声認識装置。

【請求項5】 前記利得制御手段は、前記帯域分割手段で分割された帯域信号を入力し、短時間平均パワを出力する信号パワ推定手段と、前記信号パワ推定手段からの出力信号を入力し、信号の雑音パワを推定する雑音パワ推定手段と、前記信号パワ推定手段の出力信号と前記雑音パワ推定手段の出力信号とを入力し、入力信号パワと雑音パワの割合に応じた利得量となるように前記利得制御手段の制御利得を算出する利得算出手段と、を有し、前記利得算出手段は、制御利得を a_i とし、信号パワを PXi とし、雑音パワを PNi ($i=1\sim n$)とし、利得制御範囲パラメータを Bi ($0 < Bi$)、 Di ($0 < Di$) とするとき、前記制御利得を $a_i = (PXi - Bi \times PNi + Di) / (PXi + Di)$ により算出することを特徴とする請求項1記載の音声認識装置。

【請求項6】 入力音響信号の信号 S/N を推定する信号 S/N 推定手段を更に設け、前記第1～第 n の利得制御手段の制御利得に対して、前記信号 S/N 推定手段で推定された S/N が高いときは利得の可変範囲を小さく、 S/N が低いときは利得の可変範囲を大きくすることを特徴とする請求項3～5いずれか1項記載の音声認識装置。

【請求項7】 前記信号 S/N 推定手段は、入力信号レベルの単位時間内の最大値を出力するマックスホールド手段と、入力信号レベルの単位時間内の最小値を出力するミニマムホールド手段と、前記マックスホールド手段の出力 P_{max} 及び前記ミニマムホールド手段の出力 P_{min} を入力し、2つの入力比 (P_{min} / P_{max}) 又は差 $(P_{max} - P_{min})$ に基づいて利得制御範囲パラメータ Bi を生成し、前記利得制御手段に与えるパラメータ化手段と、を有することを特徴とする請求項6記載の音声認識装置。

【請求項8】 信号 S/N 推定手段は、前記帯域分割手段で分割された各帯域毎に独立に設けたことを特徴とする請求項6又は請求項7記載の音声認識装置。

【請求項9】 前記帯域分割手段は、帯域分割数が2であることを特徴とする請求項1～8いづれか1項記載の音声認識装置。

【請求項10】 前記帯域分割手段は、クロスオーバー周波数が350Hz～500Hzの間に設定されることを特徴とする請求項9記載の音声認識装置。

【請求項11】 低域側の利得制御手段を第1の利得制御手段とし、高域側の利得制御手段を第2の利得制御手段とすると、前記第1の利得制御手段の利得制御範囲の方が、前記第2の利得制御手段の利得制御範囲より大きいことを特徴とする請求項9記載の音声認識装置。

【請求項12】 前記第1の利得制御手段が制御する利得制御範囲が15dB以下であることを特徴とする請求項11記載の音声認識装置。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】 本発明は、音声認識の耐雑音性能を向上させるために前処理装置を設けた音声認識装置に関する。

【0002】

【従来の技術】 マイクロホンを通して音声信号を音声認識装置に入力して音声認識を行うとき、高い認識率を得る条件の1つは、雑音の少ない高S/Nの音声信号を入力することである。高S/Nを得るために、通常は接話マイクロホンを、話者の口元にマイクロホンに接近させて使用する。しかし用途によっては、マイクロホンが話者の位置から数10cm～数m離れた位置に設けられる場合がある。例えばマイクロホンが機器に内蔵されていたり、アーム又はスタンドに固定されている状態で、話者自身が移動したり、話者が座っている状態で頭部を左右又は上下に向けると、口元とマイクロホンとの間隔が大幅に変動する。このような場合、マイクロホンに收音される音声信号は絶対レベルが変動すると共に、周囲騒音レベルに対して音声のレベルが相対的に低くなり、所謂S/Nが劣化する。

【0003】 一方、音声認識部は、雑音のない環境で収録された音声のスペクトルに関する辞書を持ち、照合によって認識結果を出している。従って、マイクロホンへの雑音成分が混入すると、音声認識部で音声スペクトルの照合不一致を起し、認識結果に悪影響を及ぼしてしまう。

【0004】 このような問題点の解決策として、音声認識の認識辞書側に予め雑音を想定したものを持つ方法や、マイクロホンからの出力信号を信号処理によって雑音抑圧する方法などが提案されている。ここでは、マイクロホンからの出力信号に信号処理を施し、音声認識の耐雑音性能を向上させる音声認識装置について述べる。

【0005】 先ずこのような音声認識前処理装置の構成

について説明する。図13は音声認識前処理装置を有する従来例の音声認識装置の構成図である。周波数領域変換手段1000は、入力音響信号（以下、入力信号という）を周波数領域に変換するもので、変換にはフーリエ変換手段（FFT）などが用いられる。音声検出手段1001は、入力信号が音声区間であるか否かを検出して、検出結果を出力するものである。

【0006】 雑音スペクトル推定手段1002は、周波数領域変換手段1000からの出力信号と、音声検出手段1001からの出力信号とを入力し、入力信号が音声区間でないときのパワスペクトルを平均して、平均雑音パワスペクトルを推定するものである。スペクトル減算部1003は、周波数領域変換手段1000からの出力信号と、雑音スペクトル推定手段1002からの出力信号を入力して、入力信号スペクトルから雑音スペクトルを引き去り、雑音を抑圧した音声スペクトルを合成するものである。時間領域変換手段1004は、スペクトル減算部1003からの音声スペクトルを時間領域に変換して時間波形を出力し、音声認識手段1005に出力するものである。音声認識手段1005は雑音成分の低減された音声信号を入力し、話者の発声した単語又はメッセージの内容を識別するものである。

【0007】 このような音声認識装置の動作について説明する。この処理は、一般にスペクトル減算法と呼ばれる定常的雑音の抑圧手法を用いたものであり、定常的な雑音と音声信号が混ざり合った入力信号から、音声信号のみを取り出す方式である。まず、入力信号 x は、1フレーム分（数msec～数10msec）を処理単位として周波数領域変換手段1000に入力され、信号スペクトル $X(\omega)$ に変換される。ここで $X(\omega)$ は複素数のベクトルとなる。音声検出手段1001では、例えば入力信号のレベル変化を監視していて、レベルが低く一定である部分を雑音区間と判定し、ある閾値よりレベルが大きくなる場合を音声区間などと判定して音声区間を検出する。

【0008】 雑音スペクトル推定手段1002は、音声検出手段1001が雑音区間であると判定したとき、周波数領域変換手段1000の出力 $X(\omega)$ から入力信号パワスペクトルを平均化して、定常雑音パワスペクトル $N_p(\omega)$ を推定する。スペクトル減算部1003は、入力信号スペクトル $X(\omega)$ からパワスペクトル $X_p(\omega)$ を計算し、推定音声パワスペクトル $S_p(\omega)$ を $S_p(\omega) = X_p(\omega) - N_p(\omega)$ の演算により算出する。そして入力信号スペクトル $X(\omega)$ の位相情報 $P_x(\omega)$ と、 $S_p(\omega)$ の平方根である推定音声スペクトル振幅 $S_a(\omega)$ から、推定音声スペクトル $S(\omega)$ を合成する。

【0009】 時間領域変換手段1004は、推定音声スペクトル $S(\omega)$ の時間領域信号への変換を行い、雑音成分の抑圧された音声信号を出力する。スペクトル減算

部1003において、入力信号スペクトル $X(\omega)$ から雑音パワースペクトル $N_p(\omega)$ を減算することによって雑音が除去できるのは、2つの信号が互いに独立であるとし、2つの信号が混ざり合った信号の平均パワが、夫々の信号の平均パワの加算値で近似できることに基づいている。

【0010】

【発明が解決しようとする課題】しかしながら、上記のような構成では、周波数スペクトルの導出や音声検出などの処理が必要となり、その処理量が大きくなり、コスト的にも問題が生じる。処理量を大きくしている要因として、性能を得るためにある程度の周波数分解能を必要とすることが1つの原因であり、具体的にはFFTなどの周波数領域変換手段1000や、IFFTなどで実現される時間領域変換手段1004の処理が大きくなることである。また、雑音パワースペクトル推定のために音声検出手段1001も必要となる。このための雑音抑圧処理は、音声認識手段の処理量に対して無視することのできない大きな処理量となる。例えば、サンプリング周波数を8kHzとし、周波数領域変換手段1000に64

ポイントのFFTを用いた場合、DSP(digital signal processor)を用いて処理をすると、約10～20MIPS程度の演算量を要する。

【0011】従来例では、定常的な雑音を抑圧して聴感において音声品質を向上させることにより、音声認識率の向上を期待することができる。しかし、低処理量で改善効果の大きい音声認識装置を実現することはできなかった。

【0012】本発明は、このような従来の問題点に鑑みてなされたものであって、入力信号を複数帯域に分割し、各帯域の信号を雑音レベルと信号レベルの割合に応じて適応的に利得制御し、利得制御された各帯域の出力信号を再合成し、スペクトルの全体の傾斜を補正することにより、音声認識の耐雑音性能を高めた音声認識装置を実現することを目的とする。

【0013】

【課題を解決するための手段】本願の請求項1の発明は、入力音声に混入した雑音成分によって変形した音響信号に対して、信号スペクトルの形状を補正して音声認識を行う音声認識装置であって、入力音響信号を n 個の帯域に分割する帯域分割手段と、前記帯域分割手段によって n 分割された周波数帯域の音響信号を入力し、制御利得に基づいて音響信号の増幅又は減衰を行う第1～第 n の信号増幅手段と、前記帯域分割手段からの出力信号のレベルに応じて前記制御利得を演算し、前記第1～第 n の信号増幅手段の利得を制御する第1～第 n の利得制御手段と、前記第1～第 n の信号増幅手段からの出力信号を加算し、雑音成分の低減された音声信号を出力する信号加算手段と、前記信号加算手段から出力された音声信号を入力し、音声認識を行う音声認識手段と、を具備

することを特徴とするものである。

【0014】本願の請求項2の発明は、請求項1の音声認識装置において、前記利得制御手段は、前記帯域分割手段で分割された帯域信号を入力し、短時間平均パワを出力する信号パワ推定手段と、前記信号パワ推定手段からの出力信号を入力し、信号の雑音パワを推定する雑音パワ推定手段と、前記信号パワ推定手段の出力信号と前記雑音パワ推定手段の出力信号とを入力し、入力信号パワと雑音パワの割合に応じた利得量となるように前記利得制御手段の制御利得を算出する利得算出手段と、を有し、前記利得算出手段は、制御利得を a_i とし、信号パワを PX_i ($i=1\sim n$)とし、雑音パワを PN_i とするとき、前記制御利得を $a_i = (PX_i - PN_i) / PX_i$ により算出することを特徴とするものである。

【0015】本願の請求項3の発明は、請求項1の音声認識装置において、前記利得制御手段は、前記帯域分割手段で分割された帯域信号を入力し、短時間平均パワを出力する信号パワ推定手段と、前記信号パワ推定手段からの出力信号を入力し、信号の雑音成分パワを推定する雑音パワ推定手段と、前記信号パワ推定手段の出力信号と前記雑音パワ推定手段の出力信号とを入力し、入力信号パワと雑音パワの割合に応じた利得量となるように前記利得制御手段の制御利得を算出する利得算出手段と、を有し、前記利得算出手段は、制御利得を a_i とし、信号パワを PX_i ($i=1\sim n$)とし、雑音パワを PN_i とし、利得制御範囲パラメータを B_i ($0 < B_i$) とするとき、前記制御利得を $a_i = (PX_i - B_i \times PN_i) / PX_i$ により算出することを特徴とするものである。

【0016】本願の請求項4の発明は、請求項1の音声認識装置において、前記利得制御手段は、前記帯域分割手段で分割された帯域信号を入力し、短時間平均パワを出力する信号パワ推定手段と、前記信号パワ推定手段からの出力信号を入力し、信号の雑音パワを推定する雑音パワ推定手段と、前記信号パワ推定手段の出力信号と前記雑音パワ推定手段の出力信号とを入力し、入力信号パワと雑音パワの割合に応じた利得量となるように前記利得制御手段の制御利得を算出する利得算出手段と、を有し、前記利得算出手段は、制御利得を a_i とし、信号パワを PX_i ($i=1\sim n$)とし、雑音パワを PN_i とし、利得制御範囲パラメータを B_i ($0 < B_i$)、 C_i ($0 < C_i$) とするとき、前記制御利得を $a_i = (PX_i - B_i \times PN_i) / PX_i + C_i$ により算出することを特徴とするものである。

【0017】本願の請求項5の発明は、請求項1の音声認識装置において、前記利得制御手段は、前記帯域分割手段で分割された帯域信号を入力し、短時間平均パワを出力する信号パワ推定手段と、前記信号パワ推定手段からの出力信号を入力し、信号の雑音パワを推定する雑音パワ推定手段と、前記信号パワ推定手段の出力信号と前

7
記雑音パワ推定手段の出力信号とを入力し、入力信号パワと雑音パワの割合に応じた利得量となるように前記利得制御手段の制御利得を算出する利得算出手段と、を有し、前記利得算出手段は、制御利得を a_i とし、信号パワを PX_i とし、雑音パワを PN_i ($i=1\sim n$)とし、利得制御範囲パラメータを B_i ($0<B_i$)、 D_i ($0<D_i$)とすると、前記制御利得を $a_i = (PX_i - B_i \times PN_i + D_i) / (PX_i + D_i)$ により算出することを特徴とするものである。

【0018】本願の請求項6の発明は、請求項3～5いずれか1項の音声認識装置において、入力音響信号の信号 S/N を推定する信号 S/N 推定手段を更に設け、前記第1～第 n の利得制御手段の制御利得に対して、前記信号 S/N 推定手段で推定された S/N が高いときは利得の可変範囲を小さく、 S/N が低いときは利得の可変範囲を大きくすることを特徴とするものである。

【0019】本願の請求項7の発明は、請求項6の音声認識装置において、前記信号 S/N 推定手段は、入力信号レベルの単位時間内の最大値を出力するマックスホールド手段と、入力信号レベルの単位時間内の最小値を出力するミニマムホールド手段と、前記マックスホールド手段の出力 P_{max} 及び前記ミニマムホールド手段の出力 P_{min} を入力し、2つの入力の比(P_{min}/P_{max})又は差($P_{max} - P_{min}$)に基づいて利得制御範囲パラメータ B_i を生成し、前記利得制御手段に与えるパラメータ化手段と、を有することを特徴とするものである。

【0020】本願の請求項8の発明は、請求項6又は請求項7の音声認識装置において、信号 S/N 推定手段は、前記帯域分割手段で分割された各帯域毎に独立に設けたことを特徴とするものである。

【0021】本願の請求項9の発明は、請求項1～8いずれか1項の音声認識装置において、前記帯域分割手段は、帯域分割数が2であることを特徴とするものである。

【0022】本願の請求項10の発明は、請求項9の音声認識装置において、前記帯域分割手段は、クロスオーバー周波数が350Hz～500Hzの間に設定されることを特徴とするものである。

【0023】本願の請求項11の発明は、請求項9の音声認識装置において、低域側の利得制御手段を第1の利得制御手段とし、高域側の利得制御手段を第2の利得制御手段とすると、前記第1の利得制御手段の利得制御範囲の方が、前記第2の利得制御手段の利得制御範囲より大きいことを特徴とするものである。

【0024】本願の請求項12の発明は、請求項11の音声認識装置において、前記第1の利得制御手段が制御する利得制御範囲が15dB以下であることを特徴とするものである。

【0025】

【発明の実施の形態】以下、本発明の各実施の形態の音

声認識装置について、図面を参照しつつ説明する。

【0026】(実施の形態1) 本発明の実施の形態1における音声認識装置について、図1～図7を用いて説明する。図1は本実施の形態の音声認識装置の全体構成を示すブロック図である。この音声認識装置には、入力信号を2つ以上の周波数帯に分割する帯域分割手段10が設けられる。音声信号と雑音信号とを含む信号を音響信号と呼ぶと、ここでは、マイクロホンによって入力された音響信号を2つの帯域を分割した場合について説明する。第1の帯域とは音声周波数帯の低域側を意味し、第2の帯域とは音声周波数帯の高域側を意味する。両者の境界をクロスオーバー周波数という。

【0027】先ず第1の帯域の構成要素について説明する。第1の信号パワ推定手段21は、帯域分割手段10から第1の帯域に分割された音響信号を入力して、信号パワの短時間平均を算出するものである。第1の雑音パワ推定手段31は、第1の信号パワ推定手段21からの出力信号を入力して、音声の発声されていない区間の信号、即ち雑音のみの区間のパワを推定するものである。

【0028】第1の信号パワ推定手段21からの出力信号を PX_1 とし、第1の雑音パワ推定手段31の出力信号を PN_1 とする。第1の利得算出手段41は、信号 PX_1 と、信号 PN_1 とを入力し、利得 a_1 を制御利得として決定し、第1の信号増幅手段51に与えるものである。第1の信号増幅手段51は、帯域分割手段10からの第1の帯域に分割された音響信号を入力して、所定の制御利得で増幅又は減衰して出力するものである。ここで第1の信号パワ推定手段21、第1の雑音パワ推定手段31、第1の利得算出手段41は、帯域分割手段10からの出力信号のレベルに応じて制御利得を演算し、第1の信号増幅手段51の増幅又は減衰率を制御する第1の利得制御手段を構成している。

【0029】第2の帯域の構成要素についても同様である。即ち、第2の信号パワ推定手段22は、帯域分割手段10から第2の帯域に分割された音響信号を入力して、信号パワの短時間平均を算出するものである。第2の雑音パワ推定手段32は、第2の信号パワ推定手段22からの出力信号を入力して、音声の発声されていない区間の信号、即ち雑音のみの区間のパワを推定するものである。

【0030】第2の信号パワ推定手段22からの出力信号を PX_2 とし、第2の雑音パワ推定手段32の出力信号を PN_2 とする。第2の利得算出手段42は、信号 PX_2 と、信号 PN_2 とを入力し、利得 a_2 を決定し、第2の信号増幅手段52に与えるものである。第2の信号増幅手段52は、帯域分割手段10からの第2の帯域に分割された音響信号を入力して、所定の制御利得で増幅又は減衰して出力するものである。ここでも第2の信号パワ推定手段22、第2の雑音パワ推定手段32、第2の利得算出手段42は、帯域分割手段10からの出力信

号のレベルに応じて制御利得を演算し、第2の信号増幅手段52の増幅又は減衰率を制御する第2の利得制御手段を構成している。信号加算手段60は、第1と第2の信号増幅手段51、52からの出力信号を加算して音声認識手段100に雑音成分の低減された音声信号を与えるものである。

【0031】このように構成された音声認識装置の動作について説明する。入力信号として話者からの音声信号に周囲の定常的な雑音が混入した音響信号を考える。これは実環境下でのマイクロホンに入力される信号を想定するもので、周囲騒音のレベルや、話者からマイクロホンまでの距離などによって、そのS/N値は様々に変化する。

【0032】まず、入力信号xは、帯域分割手段10に入力され、低域側と高域側の2つの周波数帯域に分割される。帯域分割手段10からの低域側の出力信号を第1の帯域信号x₁とし、高域側の出力信号を第2の帯域信号x₂とする。第1及び第2の帯域信号x₁、x₂は、夫々第1及び第2の信号パワ推定手段21、22に入力され、信号パワの短時間平均に変換される。

【0033】図2(a)は、第1及び第2の信号パワ推定手段21、22の構成例を示すブロック図である。この回路は自乗演算手段211とLPF212とにより構成される。自乗演算手段211はサンプルポイント毎に信号x_i（iは帯域番号）を自乗して出力する。そして後段のLPF212は自乗演算手段211の出力信号を平滑化し、信号パワの短時間平均値PXiを出力する。

【0034】図2(b)は、信号パワ推定手段21への入力信号x_iの波形例を示す。ここでは定常雑音、即ち音声信号のパワ変動に比較してパワ変動量が十分小さい雑音が、音声信号に重畳されている場合の波形を示す。

図2(c)は、図2(a)の構成の信号パワ推定手段21の出力する信号パワの短時間平均値（以下では信号パワという）PXiの波形例を示す。図2の(b)と

(c)とを比較して判るように、入力信号x_iのレベルに応じて信号パワPXiが大きく変動する（ただし、i＊

$$a_i = \frac{P X_i - P N_i}{P X_i} \quad \text{----- (1)}$$

ただし、i = 1, 2

$$a_i = \frac{P X_i - B_i \times P N_i}{P X_i} \quad \text{----- (2)}$$

ただし、B_iはa_iの変化範囲を制限する利得制御範囲パラメータ

※ 0 ≤ B_i, i = 1, 2

$$a_i = \left(\frac{P X_i - B_i \times P N_i}{P X_i} + C_i \right) \times \left(\frac{1}{1 + C_i} \right) \quad \text{----- (3)}$$

ただし、B_iとC_iはa_iの変化範囲を制限する利得制御範囲パラメータ

★ 0 ≤ B_i, 0 ≤ C_i, i = 1, 2

$$a_i = \frac{P X_i - B_i \times P N_i + D_i}{P X_i + D_i} \quad \text{----- (4)}$$

ただし、B_iとD_iはa_iの変化範囲を制限する利得制

御範囲パラメータ

＊ = 1, 2)。

【0035】次に、第1及び第2の信号パワ推定手段21、22からの出力信号PX₁、PX₂は、夫々第1及び第2の雑音パワ推定手段31、32に入力され、雑音パワが推定される。第1及び第2の雑音パワ推定手段の機能を低処理量で実現する方法としてミニマムホールド手段を用いるものがある。この方法では、音声検出器が無くても定常雑音のレベルを検出することができる。

【0036】図3(a)は、ミニマムホールドを実現するための演算フローを示す。先ずステップS1で信号パワPX_iを入力する。次のステップS2では、信号パワPX_iと、前回検出された雑音成分のパワ（以下では、単に雑音パワという）PN_iの大きさを比較する。PX_i > PN_iならステップS3に分岐し、PX_i ≤ PN_iならステップS4に分岐する。ステップS3では前回得られたPN_iに微小値αを加え、新たなPN_iとする。またステップS4では前回得られたPN_iを新たに入力されたPX_iの値に置き換える。こうすると、ステップS5で現フレームの雑音パワPN_iが得られる。

【0037】このようなミニマムホールド手段を用いた第1及び第2の雑音パワ推定手段31、32からの出力信号は、図3(b)のように入力信号PX_iの最低レベルを保持した雑音パワPN_iとして出力される。このようにして、PX_iはi番目の帯域の信号パワ、PN_iはi番目の帯域の雑音パワとなる。

【0038】次に、第1の利得算出手段41では、第1の信号パワ推定手段21の出力信号PX₁と、第1の雑音パワ推定手段31の出力信号PN₁とを入力して、第1の信号増幅手段51の利得a₁を算出する。同様に第2の利得算出手段42では、第2の信号パワ推定手段22の出力信号PX₂と、第2の雑音パワ推定手段32の出力信号PN₂とを入力して、第2の信号増幅手段52の利得a₂を算出する。利得a₁、a₂の決定方法は、例えば下記の(1)式、(2)式、(3)式、又は(4)式のいずれかを用いて行うことができる。

【数1】

【数2】

【数3】

【数4】

$0 \leq B_i, 0 \leq D_i, i = 1, 2$

【0039】図4(a)は、(3)式の信号処理を実現する機能ブロック図である。この場合の利得算出手段41、42は、信号増幅手段411、信号加算手段412、信号除算手段413、信号乗算手段414、信号加算手段415、信号増幅手段416で構成することができる。各信号の流れは(3)式の演算手順に準じる。即ち信号加算手段412と信号除算手段413とに PX_i が入力され、信号増幅手段411に PN_i が入力される。信号増幅手段411は信号 PN_i を、後述する利得制御範囲パラメータである増幅率 B_i で増幅し、信号 $PN_i \times B_i$ を出力する。信号加算手段412は、信号 PX_i から信号 $PN_i \times B_i$ を減算し、信号 $(PX_i - PN_i \times B_i)$ を出力する。信号除算手段413は信号 $1/|PX_i|$ を演算し、信号乗算手段414に与える。信号乗算手段414は信号加算手段412の出力と信号除算手段413の出力とを乗算し、信号 $(PX_i - PN_i \times B_i)/|PX_i|$ を出力する。信号加算手段415は信号乗算手段414の出力に利得制御範囲パラメータ C_i を加算し、 $[(PX_i - PN_i \times B_i)/|PX_i| + C_i]$ を出力する。信号増幅手段416は信号加算手段415の出力を $(1 + C_i)$ で除算し、(3)式の値を出力する。ここで、(3)式において、 $B_i = 1, C_i = 0$ としたものが(1)式であり、 $C_i = 0$ としたものが(2)式となる。

【0040】図4(b)は、図3(b)に示される信号パワ PX_i と雑音パワ PN_i とが図4(a)の利得算出手段に入力されたときの出力利得 a_i の応答波形である。図4(b)において、

波形 b_0 は、 $B_i = 1, C_i = 0$ の状態、(1)式に相当

波形 b_1 は、 $B_i = 0, C_i = 0.5$ の状態、

波形 b_2 は、 $B_i = 0.5, C_i = 0$ の状態、(2)式に相当

$$PS_i \cong PX_i - PN_i$$

よって、図6(b)の斜線部分が音声スペクトル成分となり、その下の部分が雑音スペクトル成分となる。

【0045】ここで、例えば(1)式によって利得 a_i を決定することで、信号加算手段60の出力信号は斜線部のスペクトルを取り出したものとなる。また、図6(c)は、図6(a)の時刻 t_2 における信号スペクトルを模式的に表したものであり、(b)と同様の形式で表示している。定常雑音成分は時間的に変化が少ない雑音であり、時刻 t_1 でも時刻 t_2 でも同様のスペクトルを持つが、音声信号は時々刻々と変化しているため、図6(b)から図6(c)へと変化する。上記構成は、この音声信号の時間変化に応じて利得 a_i を変化させることができ、結果として、定常雑音成分を取り除いた音声信号を得ることができる。

*波形 b_3 は、 $B_i = 0.5, C_i = 0.5$ の状態を示している。

【0041】(3)式では、 B_i をゼロに近づけるほど出力利得 a_i の変化範囲(減衰量)が制限され、 C_i が大きいほど出力 a_i の変化範囲(減衰量)が制限されることになる。

【0042】図5は(4)式の信号処理を実現する機能ブロック図である。図4(a)との違いは、信号増幅手段416が削除され、信号加算手段417が追加されたことである。この回路の信号の流れは(4)式の演算手順に準じ、構成の説明省略する。(3)式の場合と同様に利得制御範囲パラメータ D_i を大きくするほど出力利得 a_i の変化範囲(減衰量)を制限することができる。

【0043】次に、第1及び第2の利得算出手段41、42で算出された利得 a_1, a_2 に基づいて、第1及び第2の信号増幅手段51、52が信号 x_1, x_2 を夫々増幅する。信号加算手段60は $a_1 \times x_1$ と $a_2 \times x_2$ とを加算する。こうして信号加算手段60から出力された信号は、雑音成分により変化したスペクトルの傾斜やレベルが補正され、雑音成分の少ない音声信号に変換される。この音声信号は後段の音声認識手段100に入力される。音声認識手段100では、雑音による変形が取り除かれた音声のスペクトルが得られるため、音声認識率を向上させることができる。

【0044】図6(a)は定常雑音に音声信号が重畳された入力信号 x の波形を示している。図6(b)は、図6(a)の時刻 t_1 における信号スペクトルを模式的に表したもので、横軸が周波数、縦軸がパワである。今、低域側の信号パワが PX_1 、雑音パワが PN_1 、高域側の信号パワが PX_2 、雑音パワが PN_2 であるとする。定常雑音信号と音声信号が互いに独立な信号であるとすれば、短時間平均信号パワ PS_i は(5)式で近似される。

【数5】

$$PS_i \cong PX_i - PN_i$$

【0046】実際には回路規模などの問題で帯域分割数は可能な限り小さい方がよい。しかし最小の帯域分割数である2帯域では、理想的に雑音成分の影響を完全に切り除くことはできない。これは図6の(b)や(c)のスペクトルのように、雑音スペクトルや音声スペクトルが直線では表現できないことによる誤差があること、雑音パワ PN_i の推定には必ず誤差が生じること、また(5)式は近似式であることなどに起因している。誤差を持つ状態で(1)式による利得決定を行った場合、音声信号成分の欠落が生じることがあり、逆に音声認識結果を劣化させることになる。そこで、誤差の発生を予め考慮して、利得決定における利得 a_i の変化範囲に制限を設けることで、2帯域の分割数でも雑音環境下での音声認識率を改善することができる。

13

【0047】図7は、(3)式において、 $B_1 = 0.8$ 、 $B_2 = 0.8$ 、 $C_1 = 0.25$ 、 $C_2 = 0.5$ としたときの音声認識実験結果を示し、横軸を入力信号の S/N (dB)とし、縦軸を音声認識率(%)とする。この実験では、雑音を自動車走行中の雑音とし、話者は6名、認識対象用の単語数は100とした。実施の形態1の特性では、入力信号 S/N が約10dB以下の雑音状況下で、発声単語の認識性能をかなり改善できることが判る。

【0048】尚、本実施の形態では、帯域分割手段10における帯域分割数を2としたが、分割数を3以上すると、より高い効果が得られる。

【0049】(実施の形態2)次に本発明の実施の形態2における音声認識装置について、図8～図11を用いて説明する。図8は本実施の形態の音声認識装置の全体構成を示すブロック図である。この音声認識装置には、帯域分割手段10、第1及び第2の信号パワ推定手段21、22、第1及び第2の雑音パワ推定手段31、32、第1及び第2の利得算出手段41、42、第1及び第2の信号増幅手段51、52、信号加算手段60、音声認識手段100が設けられていることは、実施の形態1の場合と同様であり、それらの説明は省略する。

【0050】実施の形態1の構成要素に加えて、本実施の形態では信号 S/N 推定手段80が設けられる。信号 S/N 推定手段80は、入力信号 x の S/N を推定し、第1の利得算出手段41及び第2の利得算出手段42の利得 a_1 、 a_2 の可変範囲を、入力信号の S/N が高いときには可変範囲を小さく、入力信号の S/N が低いときには可変範囲を大きくするように制御するものである。

【0051】このような構成の音声認識装置の動作について説明する。実施の形態1では、第1及び第2の利得算出手段41、42で算出される利得 a_i の可変範囲を*

$$B = 0.5 + \frac{P_{min}}{2 \times P_{max}}$$

【0054】(6)式によれば、入力信号 x_0 の S/N が悪いとき、 P_{min} が P_{max} ほぼ等しい状態になるため、 B は1.0に近づく。逆に入力信号 x_0 の S/N が良いとき、 $P_{min} \ll P_{max}$ となり、 B は0.5に近づく。このようにして、入力信号の S/N に応じた B の値が $0.5 \leq B \leq 1.0$ の範囲で得られる。これを第1及び第2の利得算出手段41、42における(3)式の B_i に代入することで、最適の制御が行える。

【0055】このようにして、実施の形態1に対して入力信号 S/N の適用範囲が拡大し、認識率も大幅に向上する。図10は本実施の形態による音声認識の実験の結果を示す説明図である。図7と比較すると、本実施の形態では、低 S/N から高 S/N までの全 S/N 範囲に渡って認識率を更に改善されることが判る。また、本実施の形態において、帯域分割手段の帯域分割数が2のと

14

* (3)式の B_i と C_i とに固定値を与えて制限した。この方法では、認識率の改善効果はある限られた S/N の範囲になり、図7の結果の例では、 S/N が11dB以上の入力信号に対しては、逆に認識性能を劣化されてしまう。実験的に求めた結果からは、 S/N が悪いときに、利得 a_i の変化範囲を大きくした方が改善率が大きく得られ、 S/N が良いときには利得 a_i の変化範囲を小さくした方が良好な結果が得られる。これは、 S/N が悪いときには、音声認識率を劣化させている主たる要因が雑音であるため、雑音の影響を取り除く度合いを大きくした方が良い結果となることを意味する。

【0052】一方、 S/N が良いときには、雑音による音声認識率の劣化分より、処理によって利得 a_i が変動することによる波形歪みの影響が強くなる。これが高 S/N 時の認識率劣化につながる。このような理由から、本実施の形態では、信号 S/N 推定手段80において、定常雑音区間の S/N が悪い場合、(3)式の B_i の最大値を1に制限し、音声区間で雑音成分が少ないような S/N の良いときには、 B_i の最小値を0.5になるように制御する。即ち、 $0.5 \leq B_i \leq 1$ の範囲で B_i の値を適応的に変化させる。

【0053】図9は、信号 S/N 推定手段80の構成例を示すブロック図である。信号パワ推定手段801は、信号 x_0 を入力し、短時間平均パワに変換された信号 P_{X0} を出力する。ミニマムホールド手段802は、信号パワ推定手段801からの出力信号 P_{X0} を入力し、定常雑音パワレベルに対応する最小値として出力信号 P_{min} を出力する。マックスホールド手段803は、信号パワ推定手段801からの出力信号 P_{X0} を入力し、信号パワのピーク値に対応する出力信号 P_{max} を出力する。パラメータ化手段804は、例えば(6)式を用いて利得制御範囲パラメータ B を決定する。

【数6】

$$B = 0.5 + \frac{P_{min}}{2 \times P_{max}} \quad (6)$$

き、クロスオーバー周波数は、音声の重要な周波数成分が存在する200Hz～2kHzの間の値に設定するのがよい。好ましくは350～500Hzが良い。図11は、帯域分割手段10のクロスオーバー周波数を変化させた場合の音声認識率を示した説明図である。この結果から、クロスオーバー周波数は350Hz～500Hzに最適値があることが判る。

【0056】また、実施の形態1及び2において、第1及び第2の利得算出手段41、42における(3)式の定数 C_i として、低域側の定数を $C_1 = 0.25$ とし、高域側の定数 $C_2 = 0.5$ としたが、これを第1及び第2の信号増幅手段51、52での利得変化範囲に換算すると、低域側が14dB、高域側が9.5dBとなる。実験的にはこの数値が最適値となっており、信号に与える波形歪みの観点から、減衰量はこれより小さくしてお

くことが望ましい。

【0057】一般的に環境騒音は、低域成分の方がパワーが大きい。従って、スペクトル傾斜補正の観点から、低域側の補正量を大きく取る必要があり、前記のように高域側より低域側の利得変化範囲を大きくすることが望ましい。この場合、第1の利得制御手段の利得制御範囲を、第2の利得制御手段の利得制御範囲より大きくする。

【0058】尚、本実施の形態における信号S/N推定手段80での信号パワー推定手段801では、パワーでなく振幅値を求めるなど、レベルを表す出力で代用することもできる。

【0059】（実施の形態3）次に本発明の実施の形態3における音声認識装置について、図12を用いて説明する。図12は本実施の形態の音声認識装置の全体構成を示すブロック図である。この音声認識装置には、帯域分割手段10、第1及び第2の信号パワー推定手段21、22、第1及び第2の雑音パワー推定手段31、32、第1及び第2の利得算出手段41、42、第1及び第2の信号増幅手段51、52、信号加算手段60、音声認識手段100が設けられていることは、実施の形態1と同様である。

【0060】実施の形態1と異なる部分は、第1の信号帯域処理系に第1のマックスホールド手段71と第1の帯域S/Nパラメータ化手段81とを設け、第2の信号帯域処理系に第2のマックスホールド手段72と第2の帯域S/Nパラメータ化手段82とを設けたことである。

【0061】第1のマックスホールド手段71は、第1の信号パワー推定手段21からの出力信号 P_{X1} を入力し、その単位時間内の最大値 P_{max1} を出力するものである。第1の帯域S/Nパラメータ化手段81は、第1のマックスホールド手段71からの出力信号 P_{max1} と、第1の雑音パワー推定手段31からの出力信号 P_{N1} を入力し、第1の利得算出手段41に対する利得変化範囲を制限するパラメータ B_1 を出力するものである。第2の信号帯域処理系における第2のマックスホールド手段72、第2の帯域S/Nパラメータ化手段82についても、第1の信号帯域処理系のものと同一である。

【0062】このような構成の音声認識装置の動作について説明する。実施の形態3は、実施の形態2における信号S/N推定手段80に相当する機能を、帯域分割手段10からの各帯域出力毎に設けた構成としている。即ち、図9の信号パワー推定手段801を図12の第1及び第2の信号パワー推定手段21、22で置き換え、図9のミニマムホールド手段802を図12の第1及び第2の雑音パワー推定手段31、32で置き換え、図9のマックスホールド手段803を図12の第1及び第2のマックスホールド手段71、72で置き換え、図9のパラメータ化手段804を図12の第1及び第2の帯域S/Nパ

ラメータ化手段81、82で置き換えたものである。

【0063】第1及び第2の帯域S/Nパラメータ化手段81、82の動作や、第1及び第2の利得算出手段41、42の動作は、実施の形態2の場合と機能的には同様である。実施の形態3では、第1及び第2の帯域S/Nパラメータ化手段81、82によって、各周波数帯域毎に利得変化範囲の制限の調整が行えるようになる。このため利得の最適化の自由度が高まり、音声認識装置としての性能の改善が見込めるものとなる。

10 【0064】

【発明の効果】本願発明の音声認識装置によれば、周囲騒音とレベルや周波数成分に応じて入力信号のレベルや周波数特性を、音声信号成分が雑音の無い環境のものに近づくように適応的に変化させることにより、音声認識手段での認識性能を改善することができる。

【0065】特に請求項3、4、5の発明によれば、第1～第nの信号増幅手段における利得の変化範囲を制御することが可能となり、音声認識手段に対する最適化が行え、音声認識性能を改善することができる。

20 【0066】特に請求項6、7、8の発明によれば、入力信号の音声信号レベル対雑音信号レベル（S/N）の状態に応じて、第1～第nの信号増幅手段における利得の変化範囲を自動的に最適値に制御できるようになり、様々な使用環境での音声認識性能を改善することができる。

【0067】特に請求項9の発明によれば、本願の最小処理規模となる2帯域の分割で、認識率改善の効果が得られ、処理量削減の面で効果が得られる。

30 【0068】特に請求項10、11、12の発明によれば、より少ない帯域分割数で音声認識率を向上することができる。

【図面の簡単な説明】

【図1】本発明の実施の形態1における音声認識装置の全体構成を示すブロック図である。

【図2】（a）は実施の形態1の音声認識装置における信号パワー推定手段の構成例を示すブロック図、（b）は実施の形態1の音声認識装置における信号パワー推定手段への入力信号の波形図、（c）は実施の形態1の音声認識装置における信号パワー推定手段の応答例である。

40 【図3】（a）は実施の形態1の音声認識装置における雑音パワー推定手段の構成例を示すブロック図、（b）は実施の形態1の音声認識装置における雑音パワー推定手段の応答例である。

【図4】（a）は実施の形態1の音声認識装置における利得算出手段の構成例（その1）を示すブロック図、（b）は実施の形態1の音声認識装置における利得算出手段の応答例である。

【図5】実施の形態1の音声認識装置における利得算出手段の構成例（その2）を示すブロック図である。

50 【図6】（a）は実施の形態1の音声認識装置における

17

入力信号の波形例であり、(b)は入力信号の時刻 t_1 における信号スペクトル図であり、(c)は入力信号の時刻 t_2 における信号スペクトル図である。

【図7】実施の形態1の音声認識装置における音声認識率の改善効果を示す特性図である。

【図8】本発明の実施の形態2における音声認識装置の全体構成を示すブロック図である。

【図9】実施の形態2の音声認識装置における信号 S/N 推定手段の構成例を示すブロック図である。

【図10】実施の形態2の音声認識装置における音声認識率の改善効果を示す特性図である。

【図11】実施の形態1の音声認識装置における音声認識率と帯域分割手段のクロスオーバー周波数との関係を示す説明図である。

【図12】本発明の実施の形態3における音声認識装置の全体構成を示すブロック図である。

【図13】従来例の音声認識装置の全体構成を示すブロック図である。

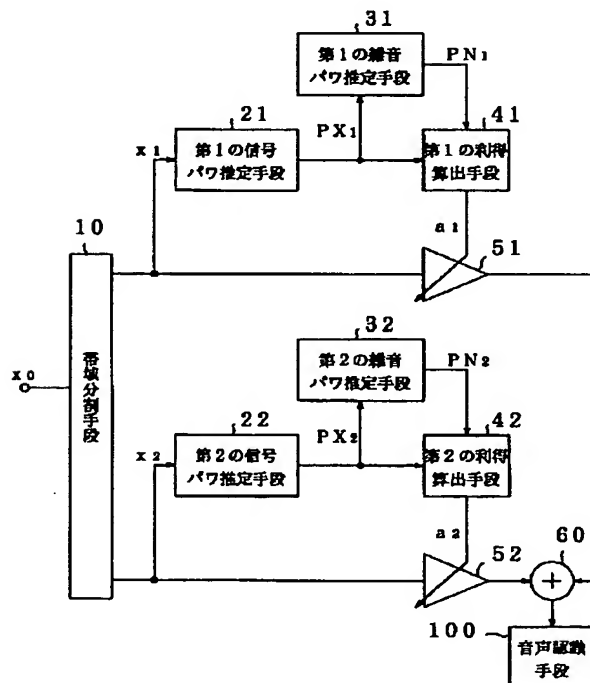
【符号の説明】

- 10 帯域分割手段
- 21 第1の信号パワ推定手段
- 22 第2の信号パワ推定手段
- 31 第1の雑音パワ推定手段

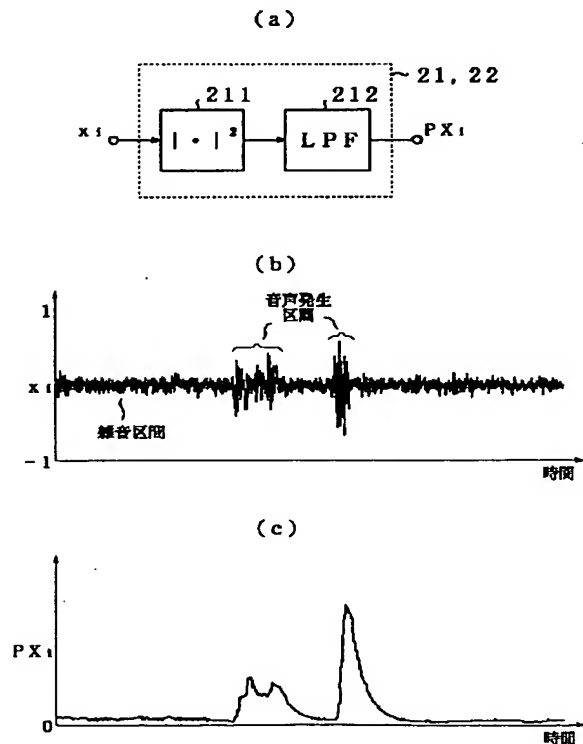
18

- 32 第2の雑音パワ推定手段
- 41 第1の利得算出手段
- 42 第2の利得算出手段
- 51 第1の信号増幅手段
- 52 第2の信号増幅手段
- 60 信号加算手段
- 71 第1のマックスホールド手段
- 72 第2のマックスホールド手段
- 80 信号 S/N 推定手段
- 81 第1の帯域 S/N パラメータ化手段
- 82 第2の帯域 S/N パラメータ化手段
- 100 音声認識手段
- 211 自乗演算手段
- 212 LPF
- 411, 416 信号増幅手段
- 412, 415 信号加算手段
- 413 信号除算手段
- 414 信号乗算手段
- 801 信号パワ推定手段
- 802 ミニマムホールド手段
- 803 マックスホールド手段
- 804 パラメータ化手段

【図1】

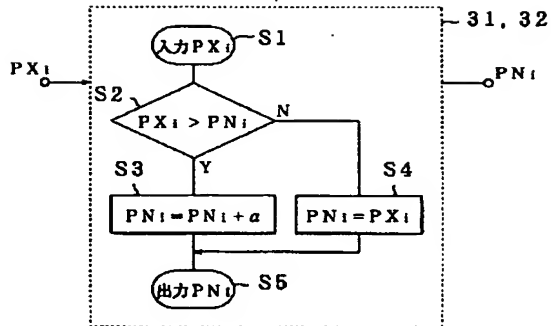


【図2】

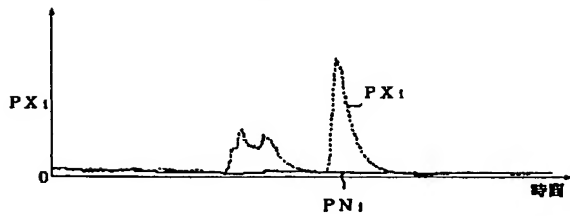


【図3】

(a)

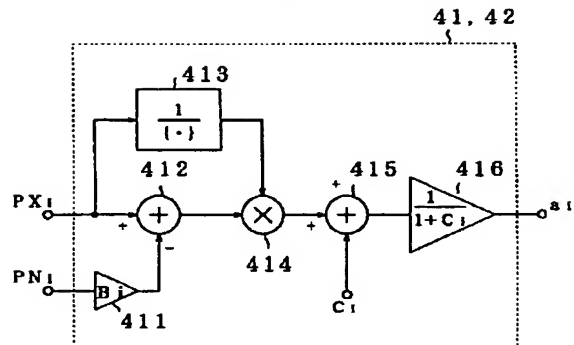


(b)

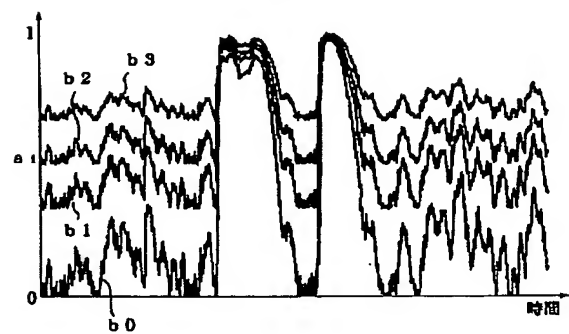


【図4】

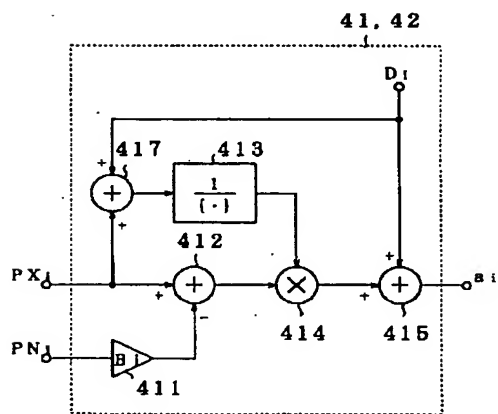
(a)



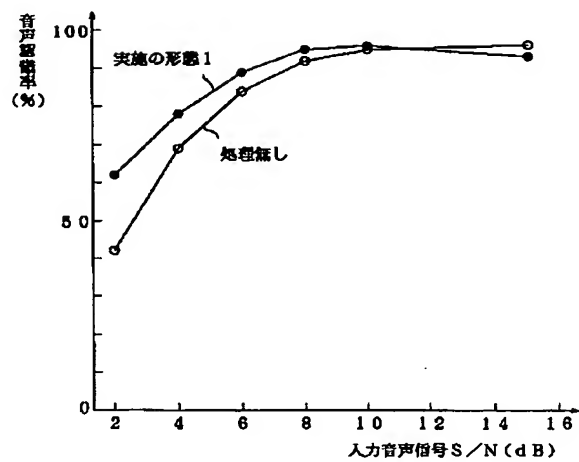
(b)



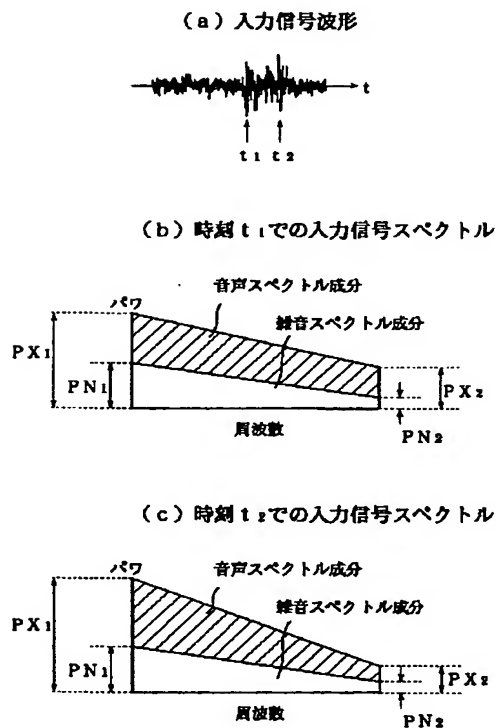
【図5】



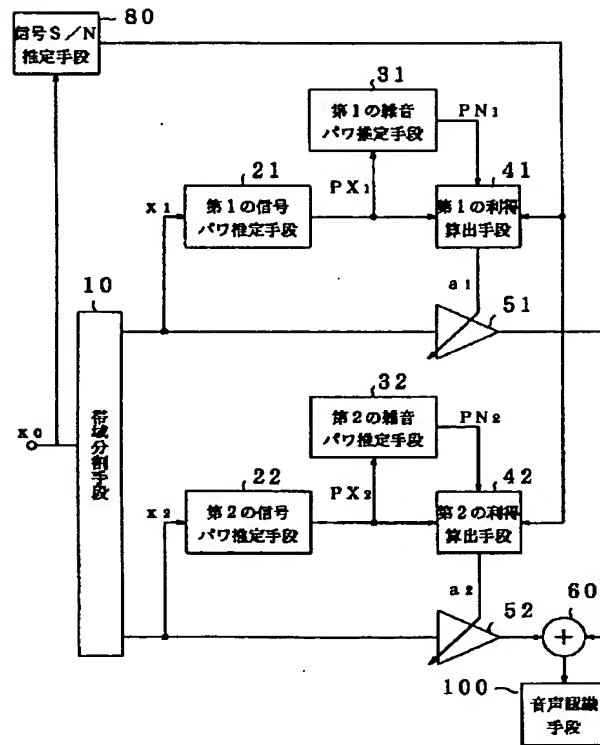
【図7】



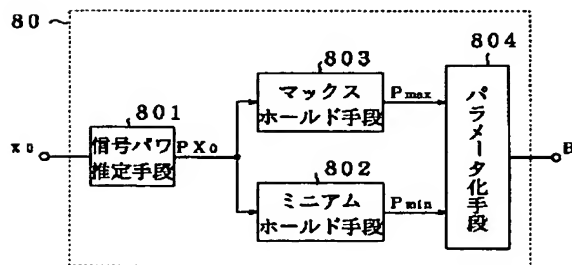
【図6】



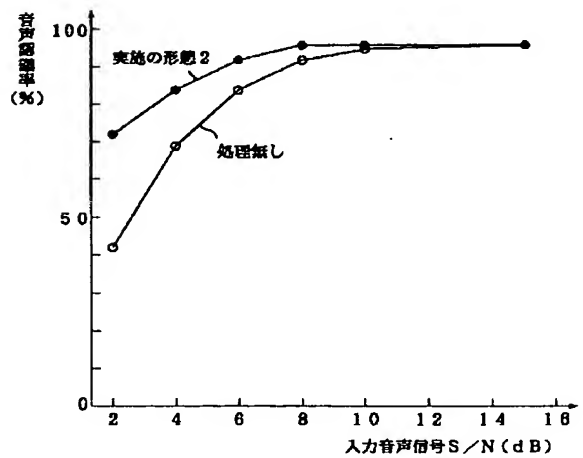
【図8】



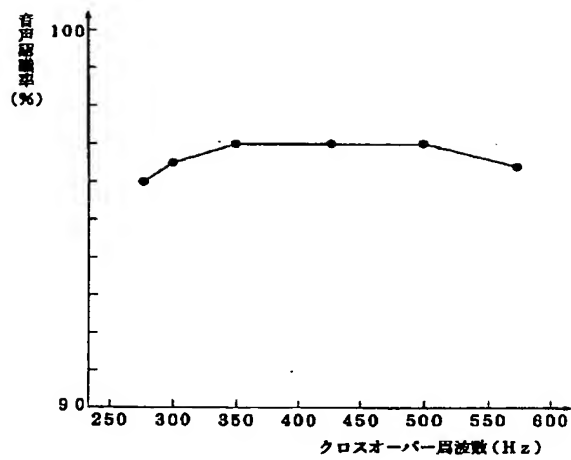
【図9】



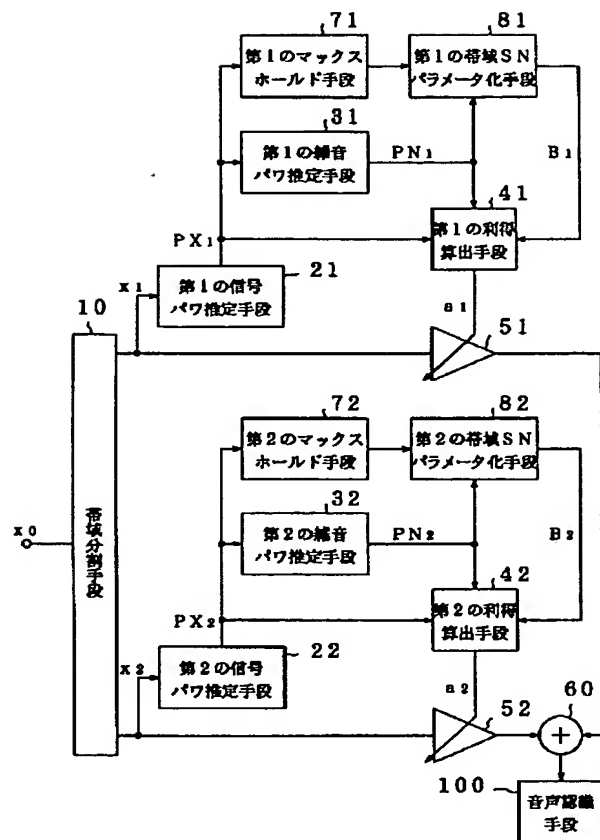
【図10】



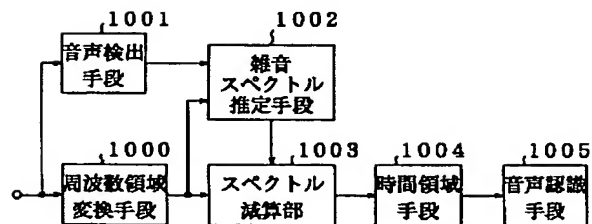
【図11】



【図12】



【図13】



フロントページの続き

(72)発明者 河村 岳
大阪府門真市大字門真1006番地 松下電器
産業株式会社内

Fターム(参考) 5D015 EE05
9A001 FF05 GG05 HH17